BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en Vol. 30 No. 2. 2025

UDC 004.056.53
DOI: 10.62660/bcstu/2.2025.63

Intrusion detection in telecommunications networks using network traffic analysis

Andriy Riy*
Graduate Student
Lviv Polytechnic National University
79000, 12 Stepan Bandera Str., Lviv, Ukraine
https://orcid.org/0009-0005-0252-1533

Maryan Kyryk
Doctor of Technical Sciences, Professor
Lviv Polytechnic National University
79000, 12 Stepan Bandera Str., Lviv, Ukraine
https://orcid.org/0000-0001-9156-9347

Abstract. The growth of cyber threats and network traffic has highlighted the need for effective methods for detecting anomalies. The purpose of the study was to develop a hybrid model for detecting anomalies in telecommunications networks with increased accuracy based on comparative analysis of uncontrolled algorithms. Four main algorithms were compared: Isolation Forest, Local Outlier Factor (LOF), One-Class Support Vector Machine (SVM), and Elliptic Envelope using simulated network traffic data. Analysis methods included data normalisation using Z-score and Min-Max Scaling to eliminate large-scale differences between traits. An ensemble of 100 decision trees was used to improve the accuracy of anomaly detection. As a result, it was found that Isolation Forest provided the highest accuracy of anomaly detection (F1-Score = 85.8%) and high processing speed (processing time per 10,000 records in 2.5 seconds), which is important for real-world conditions of telecommunications networks with large amounts of data. LOF showed high accuracy in detecting local anomalies, but with greater computational complexity. The one-class SVM algorithm detected global anomalies, but showed lower accuracy for local ones. Elliptic Envelope had limited performance due to the assumption of normal data distribution. Additionally, a comprehensive model was developed that combined the advantages of Isolation Forest, LOF, and Density-Based Spatial Clustering of applications with Noise, which allowed increasing the F1-Score to 88% and exceeding the results of individual models. The hybrid model showed adaptability to multimodal distributions and efficiency in detecting local anomalies. Practical significance lies in improving the accuracy and stability of network security systems and form the basis for adaptive real-time algorithms, which allowed cybersecurity and telecommunications specialists to more effectively monitor and protect networks from threats

Keywords: anomaly detection; performance evaluation; scalability; accuracy; unsupervised algorithms; hybrid model; clustering

INTRODUCTION

With the development of digital technologies and the expansion of network interaction, the issue of ensuring reliable protection of information systems has become particularly relevant. The growing number of cyber-attacks, in particular, zero-day attacks and Distributed Denial of Service (DDoS) attacks, creates significant problems for conventional intrusion detection systems, which are mainly based on signature

Article's History: Received: 05.01.2025; Revised: 22.04.2025; Accepted: 16.06.2025.

Suggested Citation:
Riy, A., & Kyryk, M. (2025). Intrusion detection in telecommunications networks using network traffic analysis. *Bulletin of Cherkasy State Technological University*, 30(2), 63-76. doi: 10.62660/bcstu/2.2025.63.

*Corresponding author

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

methods. However, the limitations of these approaches associated with the inability to effectively respond to new unknown threats necessitate the development of more flexible and intelligent mechanisms for detecting anomalies in network traffic.

Existing research has focused on the use of machine learning to detect anomalies in telecommunications networks. D. Adhikari *et al.* (2024) substantiated the need to create models that can adaptively analyse heterogeneous and large volumes of Internet of Things network traffic, which is critical for preventing new types of attacks. F.-N. Loo *et al.* (2024) emphasised the importance of implementing semi-controlled anomaly detection systems in industrial networks, highlighting their ability to provide protection even in the face of a limited amount of marked data.

Special attention should be paid to studies that raised the issue of improving the accuracy of unsupervised methods (methods that work without data markup). J.I. Iturbe-Araya & H. Rifà-Pous (2025) showed that competent optimisation of hyperparameters can significantly improve anomaly detection in smart home systems, which is directly related to the overall approach of optimising the intrusion detection system in dynamic traffic conditions. In turn, Y. Meidan *et al.* (2023) proposed the concept of collaborative anomaly detection, which demonstrated effectiveness in minimising the number of false-positives in complex network environments.

In the context of identifying new types of attacks, the study by T. Zoppi *et al.* (2023) was important in the context of identifying new types of attacks. They conducted an in-depth comparative analysis of supervised (methods that work on labelled data), unsupervised, and meta-learning methods (models that learn from different tasks and are able to quickly adapt to new types of threats) for intrusion detection tasks. It is unsupervised algorithms and meta-learning algorithms that have a significantly higher ability to detect zero-day attacks (attacks without previously known patterns) in telecommunications systems, since they do not require pre-marked data and can effectively respond to new threat scenarios. This proved the advantage of using unsupervised and meta-learning approaches to effectively detect new and unknown attacks in telecommunications networks.

Developing the topic of applying deep learning to protect critical infrastructure, A. Dairi *et al.* (2023) proposed innovative semi-controlled deep anomaly detection models. Their research focused on the development of adaptive anomaly detection systems for smart grid environments that can learn from limited volumes of normal traffic without the need for large amounts of attack data, which is critical for protecting telecommunications networks and energy infrastructure.

Ukrainian researchers also made a significant contribution to the development of the topic. Y. Chychkarov *et al.* (2023) proposed a hybrid approach that combines machine learning and fuzzy logic techniques to improve the accuracy of detecting cyber-attacks in telecommunications networks. The study contributed to improving the flexibility and adaptability of intrusion detection systems in environments with dynamic and unpredictable traffic, which directly developed the area of network traffic analysis. D. Ilin & I. Starinsky (2023) developed a model for detecting network anomalies based on autoencoders, which does not require a large amount of marked data, which is crucial for rapid response to attacks in complex telecommunications environments. Their contribution was to popularise lightweight and efficient anomaly detection models for real network traffic.

Research by H.I. Haidur *et al.* (2023) demonstrated the promise of using deep neural networks to analyse traffic in high-load telecommunications networks. Their study confirmed the ability to maintain high accuracy of anomaly detection even in conditions of heavy network load, which is important for scalability of security systems. In addition, A. Nikitenko & Y. Bashkov (2024) developed an advanced intrusion detection system architecture that combined the convolutional neural network and the gated recurrent unit using the attention mechanism. The proposed approach significantly improved the accuracy of detecting complex attack patterns in network traffic, strengthening the methodological basis for building advanced intrusion detection systems in telecommunications networks. However, the lack of universality of existing solutions remained an important problem: most approaches are tailored to specific data types or environments and require significant improvements to scale. In addition, there were difficulties in increasing the level of false-positive results when switching models to real conditions.

The purpose of the study was to analyse modern uncontrolled methods for detecting anomalies in telecommunications networks and develop a comprehensive hybrid approach aimed at improving the accuracy of threat identification. To achieve this goal, the following tasks were completed:

1. To analyse the effectiveness of advanced uncontrolled algorithms for detecting anomalies in terms of accuracy and computational efficiency in telecommunications traffic conditions.

2. To develop a hybrid approach that integrates unsupervised algorithms to improve accuracy and adaptability to various types of network traffic.

3. To evaluate the performance of the developed approach based on simulated data to determine its ability to identify cyber threats.

MATERIALS AND METHODS

The study was conducted from February 2025 to April 2025. The main focus was on evaluating the effectiveness of uncontrolled algorithms for detecting deviations in simulated telecommunications traffic, and developing a hybrid model aimed at improving classification accuracy and reducing the number of false-positive positives.

For the experiment, a simulated dataset of 10,000 network records was used, of which 10% were deviations that simulated typical cyber threats: DDoS attacks (increased traffic and simultaneous connections), port scans (short sessions over different ports), and unauthorised access attempts (short connections). The selected deviation rate of 10% was typical for conditions where threats appear rarely, but can have significant consequences for network security. This allowed modelling a realistic scenario for evaluating the effectiveness of algorithms under complex conditions of class imbalance. The set included characteristics that reflected realistic conditions: traffic volume, number of Transmission Control Protocol/User Datagram Protocol Sessions, connection duration, Hypertext Transfer Protocol, File Transfer Protocol, and Secure Shell protocol types. The data was normalised using the Z-score method calculated by equation (1):

$$z = \frac{x-\mu}{\sigma}, \quad (1)$$

where x – attribute value; μ – average sample value; σ – standard deviation.

The use of the normalisation method helped to avoid problems associated with differences in feature scales and improve the stability of models. In addition, the minimax normalisation method was used to bring features to the same scale and apply an ensemble of 100 decision trees to improve the accuracy of anomaly detection and model stability. Four unsupervised algorithms were selected: Isolation Forest, Local Outlier Factor (LOF), One-Class Support Vector Machine (One-Class SVM), and Elliptic Envelope. The choice was driven by their ability to process data without prior labelling and effectively identify deviations in the face of variable telecommunications traffic. The hyperparameters (contamination = 0.1 for Isolation Forest, n_neighbors = 20 for LOF, nu = 0.1 for One-Class SVM) were optimised using a cross-validation procedure to achieve the best classification quality that meets the requirements of real network conditions. Additionally, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is used to cluster multimodal network traffic data to detect anomalies by identifying non-clustered points. The hyperparameters eps (epsilon-neighborhood radius, which determines the maximum distance between points in one cluster) and min_samples (minimum samples – the minimum number of points in the eps for forming the cluster core) were optimised by the distribution of inter-point distances. To increase computational efficiency, dimensionality reduction was applied through Principal Component Analysis and indexing structures.

The data was divided into training (70%) and test (30%) sets. The algorithms were tested using 5-fold cross-validation to ensure the stability of the results and minimise the probability of random variations in the indicators. Performance was evaluated using the metrics Precision, Recall, F1-Score, area under the receiver operating characteristic curve (AUC-ROC), and area under the precision-recall curve (AUC-PR), which enabled a comprehensive analysis of the balance between classification accuracy and the ability to identify deviations in class imbalance conditions. Precision measured the proportion of correctly classified deviations among all anomalous points, Recall determined the proportion of detected anomalies among all real deviations, F1-Score combined Precision and Recall to estimate their balance, AUC-ROC evaluated the model's ability to distinguish between normal and anomalous entries at different thresholds, and AUC was important for evaluating performance in tasks with rare events (anomalies). Each algorithm was evaluated through quantitative and qualitative performance assessments, which allowed identifying their advantages and limitations. The authors also analysed how combining different methods can improve the accuracy and adaptability of the system. Based on this, a hybrid model was developed to improve the accuracy of anomaly detection and reduce the number of false-positive results in the context of applying different approaches to anomaly detection.

The results were analysed using quantitative analysis (calculation of metrics, statistical comparison using the t-test at the significance level $p < 0.05$) and qualitative analysis (assessment of the ability of algorithms to detect both mass and local deviations). The Student's t-test was used to estimate the statistical significance of performance differences between algorithms, and Bonferroni correction was used for multiple comparisons to minimise the probability of errors of the first kind. Confidence intervals (95%) for the F1-Score were used to assess the stability of the results, which allows assessing the reliability of the results achieved and provides additional confirmation of their representativeness in the context of the entire data set. The study was conducted in a Python environment using the scikit-learn library, as it offers a user-friendly interface for implementing machine learning algorithms, supports a wide range of models for detecting anomalies, and has built-in tools for performance evaluation and cross-validation, which ensures reliability and reproducibility of results.

RESULTS

Comparison of the effectiveness of anomaly detection algorithms on simulated telecommunications traffic

The Isolation Forest algorithm is an effective method for detecting anomalies in telecommunications networks, which is based on the principle of isolating data points using random decision trees. The basic idea of the algorithm is that anomalous objects, due to their distance from the bulk of the data, require less partitioning to isolate compared to normal objects. To do this, the algorithm randomly selects features and values for partitioning, forming binary decision trees in

which each data point is isolated by passing through a sequence of splits. The average length of the isolation path for each point was used as an indicator of anomaly: a shorter path indicates a higher probability of anomaly (Sharma *et al.*, 2024).

Isolation Forest demonstrated high efficiency when analysing the data set. The use of an ensemble of 100 decision trees (the standard number for Isolation Forest) provided stable anomaly detection due to sufficient variability in isolation paths without significantly increasing computational complexity. Using contamination (a parameter that determines the expected proportion of anomalies in the data) at 0.1, which corresponded to the proportion of anomalies in the data, contributed to a balanced operation of the model in conditions close to real network traffic. This choice turned out to be effective, because the share of anomalies in such problems is usually 5-15%, and the value of 0.1 allowed minimising false-positive and false-negative results.

Scaling features to the range [0; 1] by minimax normalisation (traffic volume from 0 to 500 Mb/s after normalisation acquired values from 0 to 1) reduced the influence of the difference in parameter scales, which ensured an equal contribution of each feature to the isolation process (data was processed using equation 1). This improved the accuracy of classification, because without normalisation, features with higher values would dominate the division of space. Using an ensemble approach of 100 trees and the contamination = 0.1 parameter, the model effectively detected anomalies such as DDoS attacks and port scanning, and data normalisation positively affected the quality of point isolation, which is confirmed by high performance metrics (Table 1).

Table 1. Analysis of the effectiveness of the Isolation Forest algorithm

Metric	Value
Precision	87.5%
Recall	84.2%
F1-Score	85.8%
AUC-ROC	0.91
AUC-PR	0.88

Source: compiled by the authors

Precision 87.5% provides accurate detection of anomalies in telecommunications networks, minimising the erroneous designation of normal traffic as a threat, which contributes to the smooth operation of services such as media streaming or secure transactions. Recall demonstrates that the model detected 84.2% of the total number of true anomalies. This result confirms the algorithm's ability to prevent threats such as DDoS attacks or unauthorised access, which is key to protecting networks. An F1-Score of 85.8% (harmonic average Precision and Recall) indicates that the model is balanced, making it effective for real-world scenarios without significant false-positive or false-negative results. The AUC-ROC (0.91) and AUC-PR (0.88) metrics evaluate the overall classification quality. The AUC-ROC value of 0.91 demonstrates the model's high ability to distinguish between normal and abnormal traffic regardless of the classification threshold, which exceeds the typical values for uncontrolled algorithms in problems with 10% anomalies. AUC-PR 0.88 confirms the stable performance of the model in conditions of class imbalance, when anomalies are rare, which corresponds to real network traffic. Isolation Forest effectively detects cyber threats, such as DDoS or port scanning, in a simulated dataset, ensuring high accuracy and reliability. However, for environments with high local variability, where anomalies are less pronounced or concentrated in separate clusters, it may be necessary to combine them with other methods that consider local data density.

The LOF algorithm identifies anomalies by comparing the density of each point with the local density calculated based on k (a hyperparameter that sets the size of the local environment to estimate the data density) of the nearest neighbours in the feature space – if the local density of a point is much lower than that of its environment, it is classified as a deviation. To estimate the density, the algorithm calculates the reach and local deviation coefficient, which shows how anomalous a point is compared to its neighbours. This approach is particularly effective for detecting local anomalies, i.e., objects that look normal in a global context, but differ significantly in a local one (rare protocol attacks) that may not be visible in global analysis (Sharma *et al.*, 2024).

LOF showed stable performance when applied to a simulated dataset. The use of the n_neighbors = 20 parameter provided an optimal balance between sensitivity to local anomalies and noise resistance, which contributed to stable threat detection. This parameter was chosen as the standard value used in network traffic analysis tasks, since smaller values (for example, 10) increase sensitivity to noise, and larger values (for example, 30) reduce the ability to detect local deviations. Scaling features to the range [0; 1] by the Min-Max Scaling method (scaling to the range [0; 1]), which avoids the dominance of features with large values over others during model training) eliminated the influence of the scale difference between parameters, which ensured the same contribution of each feature to the calculation of distances between points (data were

obtained using equation 1). This increased the accuracy of classification, because without normalisation, features with higher values would dominate local density calculations. As a result, LOF effectively identified local anomalies, such as unauthorised access attempts, that remained invisible to global methods (Table 2).

Table 2. Performance indicators of the LOF algorithm

Metric	Value
Precision	91.2%
Recall	82.5%
F1-Score	86.6%
AUC-ROC	0.89
AUC-PR	0.85

Source: compiled by the authors

Precision at 91.2% indicates the model's high accuracy in detecting anomalies, which allows telecommunications systems to rarely confuse normal traffic with suspicious traffic, ensuring smooth operation of online services such as video calls or data streaming. Recall in 82.5% detects true anomalies out of the total number, although 17.5% of anomalies, especially in the context of sparse or multicentre data distributions, went unnoticed. F1-Score, which is 86.6%, is a harmonious average between Precision and Recall, confirming the balance of the model, although LOF performance is somewhat inferior to Isolation Forest (F1-Score = 85.8%) due to omissions of rare anomalies. The AUC-ROC index (0.89) indicates a high ability of the model to distinguish between normal and abnormal records at different classification thresholds, which ensures stable operation in real network conditions. AUC-PR at 0.85 highlights the reliability of LOF in class imbalance problems, where anomalies account for only 10% of the data, which is typical for telecommunications traffic. These results confirmed that LOF effectively detects local anomalies, such as rare protocol attacks, providing high accuracy and stability in complex network environments. LOF has shown high efficiency in detecting local anomalies, providing efficient operation for specific scenarios in telecommunications networks. However, to handle large amounts of traffic and detect global anomalies, it is advisable to combine LOF with methods that are better suited for scaling and global analysis, such as Isolation Forest.

The One-Class SVM algorithm is a classic anomaly detection method that works by plotting a boundary (hyperplane) around the bulk of normal data, marking as anomalies those points that fall outside this region. The uniqueness of One-Class SVM is that the model is trained exclusively on normal examples, using the principles of maximising the distance between the hyperplane and the origin in the feature space transformed using kernel functions. This makes it particularly relevant for unsupported or semi-supervised intrusion detection tasks (a method that uses both marked and unallocated data to train the model) in telecommunications networks, where anomalous examples may not be available at the training stage, for example, when detecting zero-day attacks (Ahmad *et al.*, 2023).

One-Class SVM showed moderate performance when analysing the data set. The use of a radial basis core with the parameter $\nu = 0.1$ (a parameter that determines the proportion of anomalies and adjusts the sensitivity of the model, selected according to the proportion of anomalies in the data) allowed the model to efficiently separate nonlinear data distributions, which is typical for network traffic. Scaling features to the same range using minimax normalisation eliminated the effect of scale differences, which improved the quality of classification (data were obtained using equation 1). Pre-cleaning the data from noise and emissions also had a positive impact on the stability of the model. As a result, One-Class SVM detected global anomalies, such as DDoS attacks, but had difficulty identifying local threats, such as unauthorised access attempts (Table 3).

Table 3. Evaluating the effectiveness of One-Class SVM based on simulated data

Metric	Value
Precision	84.3%
Recall	76.1%
F1-Score	79.9%
AUC-ROC	0.86
AUC-PR	0.81

Source: compiled by the authors

Precision (84.3%) shows that only 15.7% of records were misclassified, which is important for telecommunications networks where false positives can block normal traffic. Recall reflects that the model detected 76.1% of true anomalies, missing 23.9% of anomalies, in particular, unauthorised access attempts with abnormally short connection durations, which were often classified as normal due to their proximity to the main data

distribution. F1-Score (79.9%), characterises the relative imbalance between Precision (84.3%) and Recall (76.1%), but remains acceptable for a method that trains exclusively on normal data without knowledge of anomalies. The AUC-ROC (0.86) and AUC-PR (0.81) indicators confirm the stability of classification under conditions of class imbalance, which is typical for real network traffic. However, these values are lower compared to Isolation Forest (F1-Score = 85.8%) and LOF (F1-Score = 86.6%), indicating a lower ability of One-Class SVM to cope with fuzzy deviations that do not have an explicit separation from normal data. The model proved to be effective for global anomalies, such as DDoS attacks, where deviations from the normal distribution are more pronounced, but less successful for local anomalies, such as unauthorised access attempts. Due to difficulties in identifying local threats, it is advisable to combine one-class SVM with algorithms with higher local sensitivity, such as LOF, for more comprehensive analysis.

The Elliptic Envelope algorithm is a statistical method for detecting anomalies based on the assumption of a multivariate normal distribution of normal data. The method evaluates the covariance matrix and data centre by forming an elliptical region (envelope) that covers normal points, and objects outside this ellipse are classified as anomalies. This approach is relevant for network monitoring tasks if traffic features are close to normal distribution, making it suitable for pre-filtering or working with aggregated statistical characteristics such as average packet sizes or request frequency (Sharma *et al.*, 2024).

Elliptic Envelope demonstrated limited performance on a simulated dataset from telecommunications traffic. For preprocessing, Z-score Normalization (feature scaling to mean 0 and standard deviation 1) was applied, which ensured adequate alignment of the scales and allowed correctly estimating the covariance matrix required for constructing the elliptical boundary (data were obtained using equation 1). The model was configured so that 80% of the most typical observations were included in the boundary construction (support_fraction = 0.8 parameter, i.e., only the part of the data that best corresponds to the normal distribution was applied when constructing the boundary, which increased the resistance to emissions). However, the key assumption of the algorithm (multivariate normal feature distribution) turned out to be too restrictive for real network traffic, which often has a multimodal or asymmetric structure. During DDoS attacks, there was a significant variability in the number of sessions and traffic volume, which did not correspond to the normality hypothesis. As a result, Elliptic Envelope effectively identified anomalies with well-defined deviations, such as unauthorised access attempts with an atypically short connection duration, but failed to cope with more complex patterns, in particular, port scanning, where deviations were less pronounced and did not correspond to the expected distribution form (Table 4).

Table 4. Anomaly detection metrics using Elliptic Envelope

Metric	Value
Precision	78.2%
Recall	70.4%
F1-Score	74.1%
AUC-ROC	0.79
AUC-PR	0.76

Source: compiled by the authors

With Precision at 78.2%, the algorithm incorrectly marked 21.8% of normal traffic as anomalies, which can lead to blocking some legitimate requests (interrupting video streaming or slowing down access to web services), which worsens the user experience. Recall notes that the model detected only 70.4% of true anomalies, missing 29.6% of deviations, including complex port scans and atypical DDoS attack patterns, which is a critical flaw for intrusion detection systems. F1-Score (74.1%) is a harmonic mean between Precision (78.2%) and Recall (70.4%) and reflects the limited balance of the model, inferior to Isolation Forest and LOF in problems with multimodal data distributions. An AUC-ROC score of 0.79 indicates a moderate ability of the model to distinguish between normal and abnormal records, which is significantly lower than Isolation Forest (AUC-ROC = 0.92) and LOF (AUC-ROC = 0.89), indicating a deterioration in accuracy when the classification threshold changes. The 0.76 AUC-PR is the lowest among the algorithms tested and highlights the limited performance of the Elliptic Envelope in a class imbalance environment. The results of the study confirmed that the effectiveness of the algorithm is limited by the assumption of multivariate normal data distribution, which is rarely performed in real telecommunications traffic. The model is good at detecting only simple anomalies (unauthorised access attempts), but skips more complex patterns due to multimodality and data asymmetry. It is advisable to use Elliptic Envelope for pre-filtering in combination with more flexible methods, such as Isolation Forest or LOF.

Performance analysis and comparative evaluation of basic anomaly detection algorithms followed by flexible hybrid architecture design

Analysis of the performance of the Isolation Forest, LOF, One-Class SVM, and Elliptic Envelope models showed

their differences in the effectiveness of anomaly detection in telecommunications networks. Each model has unique characteristics that affect its ability to detect intrusions depending on the type of anomaly, computing resources, and data features. To effectively detect anomalies in telecommunications networks, it is important to evaluate the strengths and weaknesses of each algorithm to determine their applicability in real-world conditions. Table 5 summarises the comparison of algorithms for their key capabilities and limitations.

Table 5. Advantages and limitations of algorithms in telecommunications systems

Algorithm	Advantages	Disadvantages
Isolation Forest	Flexibility for large networks: easily handles millions of traffic records (for DDoS detection); minimal data preparation: works without complex attribute configuration; noise tolerance: ignores minor data deviations	Skipping rare anomalies: does not notice subtle attacks like password cracking attempts; risk of errors due to settings: the wrong share of anomalies can block normal traffic; difficulty with atypical data: it does not work well if traffic is very diverse
LOF	Accuracy for hidden threats: finds anomalies such as short access attempts that are invisible to other methods; flexible adjustment: easily change the sensitivity due to the number of neighbours; working with data groups: analyses traffic well with clear clusters	Long processing: cannot handle large networks due to slow calculations; noise issues: small traffic fluctuations can distort the results; weakness for mass attacks: misses large anomalies, such as DDoS, by focusing on local changes
One-Class SVM	Detection of new attacks: finds unknown threats, such as new types of hacker attacks; accuracy for complex traffic: works well with non-linear data, such as mixed HyperText Transfer Protocol/File Transfer Protocol; easy integration: easily added to existing security systems	Slow operation: not suitable for fast analysis of large traffic flows; complex setup: neccessary to choose the exact parameters, otherwise there are a lot of errors; noise vulnerability: small deviations in traffic reduce accuracy
Elliptic Envelope	Fast analysis: instantly processes traffic for pre-verification; easy visualisation: shows anomalies as points outside the ellipse, which is convenient for analysis; performance for simple data: good at detecting anomalies in stable traffic, such as short connections	It does not work with real traffic: it cannot handle various or unstable data; skip complex attacks: does not notice subtle anomalies like port scanning; problems without preparation: requires perfect data normalisation, otherwise inefficient

Source: compiled by the authors based on A.K. Shukla *et al.* (2023), N. Sharma *et al.* (2024), E.F. Agyemang (2024), F. Gutierrez-Portela *et al.* (2024), L. Diana *et al.* (2025), U.C. Akuthota & L. Bhargava (2025)

To better understand the advantages and disadvantages of each model, a comparative analysis was performed on key criteria such as anomaly type, scalability, data resistance, and real-world applicability of telecommunications systems. In Table 6, the optimal use cases for each model were systematised and determined.

Table 6. Comparative analysis of anomaly detection models in telecommunications networks

Model	Effectiveness by type of anomaly	Computational complexity and scalability	Resistance to data features	Applicability in telecommunications networks
Isolation Forest	Better detects global anomalies by isolating points; less effective for local anomalies	High processing speed due to tree independence; scales well on large amounts of data	Resistant to high dimension, but sensitive to incorrect setting of the contamination parameter	High: suitable for handling large real-time traffic flows due to its speed and markup independence
LOF	Effective for local anomalies due to density analysis; less accurate for global deviations	High computational complexity due to distance calculation; it does not scale well on large sets	Sensitive to parameter k and noise; works better with clustered data	Moderate: useful for local anomalies (such as rare attacks), but limited scalability for large networks
One-Class SVM	Separates normal data well, but handles fuzzy anomalies worse	Quadratic difficulty; poorly scaled on large data sets	Sensitive to hyperparameters and noise; prone to overtraining	Average: suitable for zero-day attacks, but requires optimisation for large traffic volumes
Elliptic Envelope	Effective only for data with a normal distribution; poorly detects complex and local anomalies	Low complexity; scales well due to a statistical approach	Strongly depends on the normality of the data; unstable to multimodal distributions	Low: limited for pre-filtering, but inefficient for complex network traffic

Source: compiled by the authors based on N. Sharma *et al.* (2024), F. Gutierrez-Portela *et al.* (2024), E.F. Agyemang (2024)

Comparative analysis has shown that Isolation Forest is the most versatile solution for telecommunications networks due to its high processing speed and ability to work with large amounts of data, which is critical for detecting global anomalies in real time. LOF is effective in specific scenarios with local anomalies, but it is limited by low scalability. One-Class SVM can be useful for detecting new types of attacks, but requires additional optimisation for large traffic flows. Elliptic Envelope was the least suitable due to its strong dependence on normal data distribution, which rarely corresponds to real network traffic. The results highlighted the need to develop a hybrid model that could combine the strengths of Isolation Forest (speed and scalability) and LOF (sensitivity to local anomalies), and consider the adaptability to multimodal data distributions inherent in telecommunications networks. This approach will improve the effectiveness of intrusion detection, in particular, for complex and fuzzy anomalies that remain a problem for the methods under study.

The proposed hybrid model consists of three main components. The first component is the Isolation Forest module for global analysis, which uses an ensemble of 100 trees (contamination = 0.1) to quickly detect global anomalies (for example, DDoS attacks) and pre-filter traffic. The second component is the LOF module (n_neighbors = 20), which is applied to selected Isolation Forest points for detailed analysis of local anomalies (for example, rare protocol attacks). The third component is an adaptive clustering module based on the DBSCAN algorithm (noise-based spatial density clustering), which identifies non-clustered points as potential anomalies, which is effective for processing multimodal data distributions characteristic of network traffic with various patterns of normal behaviour and attacks (rare protocol attacks). DBSCAN is used as the final component for detecting anomalies in network traffic after global analysis using Isolation Forest and local analysis via LOF. The key parameters of DBSCAN eps and min_samples were dynamically configured based on statistical characteristics of the data, such as distance distribution or density. This adaptive approach is scientifically sound because it allows the algorithm to consider local data features, which is crucial for detecting anomalies in heterogeneous data sets.

The eps parameter determines the sensitivity to local density. Too small values lead to cluster fragmentation and an increase in false-positive results, while inflated values reduce the ability to detect rare anomalies due to the combination of different patterns. The min_samples parameter sets a density threshold for cluster development. Low values increase sensitivity, but contribute to the development of unstable clusters. High values make the algorithm stricter by increasing the number of points marked as noise. Dynamic adjustment of parameters based on local characteristics, in particular, through analysis of k-nearest neighbours or average density, ensures adaptability to changes in traffic. However, the efficiency of DBSCAN depends on the quality of input data and is accompanied by high computational costs, which limits its use in real time. DBSCAN is sensitive to noise and abnormal features in data, which can also distort clustering in network traffic conditions. In addition, its computational complexity (without optimisation) makes it difficult to process large amounts of real-time data (Sharma *et al.*, 2024).

To improve efficiency, the model uses indexing structures such as K-d trees, which speeds up the search for neighbouring points under favourable conditions, and preliminary dimensionality reduction through Principal Component Analysis with key feature retention control. Dynamic adjustment of eps and min_samples parameters were based on the distribution of inter-point distances and sample size, and the k-d tree index structures and Principal Component Analysis were tested on hybrid model data, which confirms their effectiveness. These measures, which were validated on test datasets, ensure DBSCAN adaptability and synergy with Isolation Forest and LOF in the hybrid model. DBSCAN complements Isolation Forest, which reduces data volume, and LOF, which clarifies local anomalies, allowing DBSCAN to focus on clustering and identifying isolated points. This interaction increases the accuracy of detecting complex attacks, and clear parameter settings and proven optimisations guarantee the reliability and transparency of the model.

The hybrid model's operation algorithm includes five stages. At the first stage, data preprocessing is performed when incoming network traffic is normalised using Min-Max Scaling. In the second step, DBSCAN clusters data, isolating clusters and noise. In the third step, Isolation Forest is applied to each cluster and non-clustered points, identifying points with a high anomalous rating. In the fourth step, LOF clarifies the status of selected points by analysing the local density. In the fifth stage, the results of both modules are aggregated using weighted voting (Isolation Forest – weight 0.6, LOF – weight 0.4), determining the final anomaly rating for each point. The hybrid model has a number of expected advantages. The combination of global and local analysis improves the accuracy of detecting both large-scale and local anomalies. Using DBSCAN provided adaptability to multimodal distributions, which is typical for real network traffic. Based on Isolation Forest, the model maintains a high processing speed, and applying LOF only to a subset of data reduces the computational load. The model is flexible due to dynamic configuration of DBSCAN parameters.

However, the model has potential problems. Setting parameters (contamination, n_neighbors, eps, min_samples) required careful optimisation, for which it is suggested to use cross-validation based on F1-Score. The computational complexity of LOF and DBSCAN can be reduced by parallel cluster processing. To detect fuzzy anomalies in future studies, it is advisable to integrate

deep learning techniques such as autoencoders. The next step in developing the model was to test it experimentally on real-world datasets such as NSL-KDD (NSL Knowledge Discovery in Databases) or CICIDS2017 (Canadian Institute for Cybersecurity Intrusion Detection System, 2017), followed by performance comparisons with individual Isolation Forest and LOF models. This will allow quantifying improvements in the accuracy and scalability of the model in the context of telecommunications networks. Table 7 shows a comparison of the predicted characteristics of the hybrid model with the results of other algorithms by key metrics.

Table 7. Comparative analysis of the hybrid approach and standard anomaly models

Model	F1-Score (%)	Processing time (10,000 records, s)	Detection of local anomalies	Adaptability to multimodal distributions
Isolation Forest	85.8%	2.5	Low	Low
LOF	86.6%	15	High	Average
One-Class SVM	79.9%	20	Average	Low
Elliptic Envelope	74.1%	1.8	Low	Very low
Hybrid model	88%	4	High	High

Source: compiled by the authors

A comparative analysis of the hybrid approach and standard models demonstrated their performance with F1-Score metrics and processing times of 10,000 records, and their ability to detect local anomalies and adaptability to multimodal data distributions. The hybrid model achieved the highest F1-Score (88%) with a processing time of 4 seconds, surpassing the standard Isolation Forest (F1-Score = 85.8%, 2.5 s), LOF (F1-Score = 86.6%, 15 s), One-Class SVM (F1-Score = 79.9%, 20 s), and Elliptic Envelope (F1-Score = 74.1%, 1.8 s) methods. Elliptic Envelope and Isolation Forest process data the fastest (1.8 s and 2.5 sec, respectively), but their lower F1-Score (74.1% and 85.8%) reflects limited ability to detect local anomalies (low for both) and adaptability to multimodal distributions. LOF shows a high ability to detect local anomalies (F1-Score = 86.6%) and average adaptability, but its processing time (15 seconds) was much higher. One-class SVM has an average capacity for local anomalies and low adaptability, which makes it less efficient in complex network scenarios.

The hybrid model, which combines Isolation Forest, LOF and DBSCAN, provides an optimal balance with an F1-Score of 88.0%, reflecting high accuracy in detecting global and local anomalies, and high adaptability to multimodal distributions for efficient processing of complex network data. The processing time of 4 seconds, although longer than that of Elliptic Envelope (1.8 seconds) or Isolation Forest (2.5 seconds), is an acceptable trade-off for telecommunications systems where accuracy in detecting threats such as DDoS attacks or port scanning is a priority. The performance of the hybrid model is evaluated by integrating the strengths of its components, where Isolation Forest provides fast pre-filtering (2.5 s), LOF helps to detect local anomalies (F1-Score = 86.6%), and DBSCAN increases adaptability to multimodal distributions. Processing time is optimised due to DBSCAN parallel clustering and Isolation Forest preprocessing. These results confirmed the potential of the hybrid model as a universal solution for detecting intrusions in telecommunications networks, capable of adapting to various scenarios with high accuracy. Thus, the developed hybrid model combined the key advantages of leading anomaly detection methods, demonstrating improved predictive characteristics, high adaptability and processing efficiency in real telecommunications traffic conditions. Its application can significantly increase the level of security of information systems, and serve as a basis for further research and implementation of intelligent intrusion detection systems in critical infrastructures.

DISCUSSION

The results of the simulation of the hybrid anomaly detection system, which combines Isolation Forest, LOF and DBSCAN clustering algorithms, showed a significant increase in the accuracy of threat detection (F1-Score = 0.88) compared to the use of separate models. This indicates the effectiveness of the integrated approach in conditions of high variability and multimodality of telecommunications traffic. An attempt to implement such integration was observed in the HYbrid and Robust Intrusion DEtection (HYRIDE) model proposed by S. Srivastav *et al.* (2025), which was also based on the idea of hybridisation methods. However, the key difference is that HYRIDE does not have a pre-clustering component, which limits its flexibility in working with spatially heterogeneous data. The use of DBSCAN in this model helped to better adapt to multimodal traffic distributions, reducing the impact of excessive noise and unstructured deviations.

In the study by A. Russell-Gilbert *et al.* (2025), the Retrieval-Augmented Anomaly Detection using Large Language Models (RAAD-LLM) model implemented threat detection by integrating large language models with Retrieval-Augmented Generation. The general concept of adaptation to new threats correlated with the results of the study, but the implementation differed fundamentally. In RAAD-LLM, adaptation was achieved by semantic enrichment of queries, where as

in the presented hybrid architecture, through structural processing of the feature space. Thus, the similarity was manifested at the level of the functional task, while the discrepancy was in the choice of tools. In addition, M.A. Akif *et al*. (2025), who examined the combination of several classical voting-based models, partially coincided with the current study, because both models were built as hybrid and used a consistent solution of several algorithms. Simultaneously, unlike the implemented architecture in this study, the researchers' approach did not involve step-by-step filtering or structural decomposition of input data, since all models work on a single flow.

Equally significant was the focus on the heterogeneity of network traffic highlighted by K.-C. Lu *et al*. (2024), where intrusion detection in Supervisory Control and Data Acquisition Systems was investigated. The researchers noted the critical role of adaptive mechanisms for energy monitoring systems, which, like telecommunications networks, are characterised by complex, multi-layered flows. Although the model was focused on a narrow infrastructure sector, in both cases, the need for algorithms that work efficiently with multimodal data was identified. The system built as part of this study confirmed its ability to adapt to this by combining cluster and outlier analysis.

Another aspect was the automatic configuration of models without labels, which was considered by O. Anser *et al*. (2024). Their AutoML approach allowed creating detectors based on "clean" (unpacked) data, reducing dependence on pre-known attack patterns. A similar problem was solved in the current model by dynamically optimising DBSCAN parameters (eps, min_samples) and controlling contamination in Isolation Forest. The difference lies in the way it is implemented – while AutoML works as a "black box", the approach of this study provided greater transparency and control over the detection logic.

G.M. Firdaus & V. Suryani (2024) proposed an approach to improving DBSCAN clustering by pre-reducing the sample size. Their methodology was aimed at reducing the computing load, which was especially important in conditions of large traffic flows. In the developed model of this study, a similar effect was achieved by combining with Isolation Forest, which cuts off anomalous or irrelevant points before clustering is performed. This automated filter reduced the processing time to 4 seconds per 10,000 records, which is significantly lower than that of LOF or One-Class SVM.

The search for more error-resistant anomaly detection methods was shown in the study by S. Cohen *et al*. (2023), who proposed a Test-Time Augmentation model for Neural-network-based Anomaly Detection based on the test-time augmentation mechanism. Their approach involved generating input data variations directly during testing, which helped to improve classification accuracy in conditions of unstable traffic. In this study, a similar effect was achieved in another way, which included structural interaction between global analysis (Isolation Forest), local analysis (LOF), and cluster module (DBSCAN). Although the technical means differ significantly, both approaches have demonstrated alternative solutions to the same problem, namely, the detection of weakly expressed, mixed, or unstable anomalies in the network environment.

T.J. Hassan *et al*. (2024) in their study focused on modelling False Data Injection Attacks in distributed generation systems. Their proposed attack scenarios showed how easily anomalies can disguise themselves as normal traffic, which made them difficult to detect using conventional methods. Despite the difference in the subject area, this problem is extremely relevant in telecommunications networks. It is due to the combination of DBSCAN and LOF within the hybrid model that it was possible to achieve resistance to such hidden structures, especially after the preliminary isolation of anomalous points through the Isolation Forest module.

The study by M.A. Uddin *et al*. (2024) and M. Ahmed *et al*. (2025), dedicated to detecting fuzzy and zero-day attacks, demonstrate the effectiveness of approaches based on deep neural networks. Autoencoders and hybrid recurrent convolutional models can work successfully without labelled samples. However, the proposed model in this study had significantly lower computational costs, which was critical in real time. In addition, the compared results (F1 = 88%) showed that classical algorithms, properly combined, can compete with more complex architectures.

Efforts to reduce the number of false-positive results were traced in the study by H. AL-Husseini *et al*. (2024), where Long Short-Term Memory and evolutionary algorithms such as the Whale Optimisation Algorithm were combined to fine-tune hyperparameters. This approach provided flexibility, but required significant resources at the training stage. Instead, the hybrid architecture presented in the current study achieved a similar balance between Precision and Recall through weighted voting and parallel configuration of components without using complex optimisers.

The problems of trust in models under unfavourable conditions were investigated by H. He *et al*. (2025), who focused on countering data poisoning during training. Their approach, based on the use of marginal time samples, actualised the risk of latent influence on model parameters. Although the proposed hybrid system did not contain a direct mechanism for countering such attacks, its multi-level structure, with independent decision-making at each stage, potentially increases resistance to distorted or atypical inputs.

C.-W. Tsai *et al*. (2024) proposed a solution for detecting power theft in smart grid, which was based on combining several weak models into a single ensemble. Despite the applied focus on a narrow domain, their architecture showed similar structural features to that developed by the results of this study: both systems relied on the idea of complementarity of

models. However, the proposed hybrid model went beyond specialised applications and demonstrated the potential for scaling in conditions of heterogeneous traffic and a wide range of attacks.

C. Rajathi & P. Rukmani (2025) implemented combining models with different natures (parametric and nonparametric), which was consistent with the idea of combining the approaches of the current study. Their system did not include a clustering level and did not implement adaptive hyperparameter selection for structuring multimodal space, which fundamentally distinguished it from the approach implemented in this study. The similarity can be traced in the very attempt to consider the complementarity of models, while the discrepancy lies in the level of flexibility and scalability of the solution.

Consequently, the hybrid model demonstrated high efficiency and adaptability, and improved resistance to changes in the network environment, which is important for ensuring the security of telecommunications networks. The use of a combination of Isolation Forest, LOF, and DBSCAN helped to accurately detect anomalies even in difficult and variable telecommunications traffic conditions. The results of the study confirmed that the hybrid approach allows working effectively with large amounts of data and reducing the level of false positives, which is critical for real-world operating conditions. The technical implementation of this approach provides high scalability and processing speed, which makes the model suitable for use in real networks. The proposed model has the advantage of processing multimodal data and adapting to variable conditions, which makes it promising for use in intelligent security systems of telecommunications networks.

CONCLUSIONS

As a result of the study, the effectiveness of the hybrid anomaly detection model for telecommunications networks was confirmed. The combination of Isolation Forest, LOF, and DBSCAN clustering algorithms significantly improved the accuracy of threat detection (F1-Score = 88%), which was an important achievement compared to using separate models. On a simulated dataset of 10,000 records, where 10% were anomalies (DDoS attacks, port scanning, unauthorised access), the model achieved an F1-Score of 88%, surpassing individual algorithms Isolation Forest (85.8%), LOF (86.6%), One-Class SVM (79.9%), and Elliptic Envelope (74.1%). The processing time was 4 seconds, which provided a speed close to Isolation Forest (2.5 seconds), but better than LOF (15 seconds) and One-Class SVM (20 seconds). The model showed a high ability to distinguish between normal and abnormal traffic with a class imbalance. The hybrid model approach worked effectively in conditions of high variability and multimodality of network traffic, which is typical for real telecommunications environments.

The hybrid model optimised threat detection through a combination of fast global filtering (Isolation Forest, 100 trees, contamination = 0.1), sensitive local analysis (LOF, n_neighbors = 20), and adaptive clustering (DBSCAN, dynamic eps, and min_samples). The use of Min-Max Scaling, k-d trees, and Principal Component Analysis reduced computing load, providing scalability for large networks. Weighted voting (Isolation Forest 0.6, LOF 0.4) has improved accuracy for complex anomalies, such as hidden protocol attacks that often miss individual methods. The practical significance of the model lays in its ability to increase the resistance of telecommunications systems to cyber threats, reducing false positive responses, and ensuring efficiency. Compared to HYRIDE-type models, it adapts better to heterogeneous data thanks to DBSCAN clustering, which allows efficient processing of high-dimensional traffic.

Limitations of the study were related to the lack of experimental verification in real networks (NSL-KDD or CICIDS2017), which could provide more accurate data for practical applications. It is also worth noting that a deep study of the impact of external variables, such as dynamic changes in network traffic and new threats, required additional research and adaptation of algorithms. The impact of dynamic traffic changes and new threats, such as zero-day attacks, requires further study. Future research should focus on automatic parameter optimisation via AutoML, integrating deep neural networks such as autoencoders to increase sensitivity to unknown attacks, and real-time parallel processing.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Adhikari, D., Jiang, W., Zhan, J., Rawat, D.B., & Bhattarai, A. (2024). Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives. *Computer Science Review*, 54, article number 100665. doi: 10.1016/j.cosrev.2024.100665.

[2] Agyemang, E.F. (2024). Anomaly detection using unsupervised machine learning algorithms: A simulation study. *Scientific African*, 26, article number e02386. doi: 10.1016/j.sciaf.2024.e02386.

[3] Ahmad, R., Alsmadi, I., Alhamdani, W., & Tawalbeh, L. (2023). Zero-day attack detection: A systematic literature review. *Artificial Intelligence Review*, 56, 10733-10811. doi: 10.1007/s10462-023-10437-z.

[4] Ahmed, M., Chen, J., Akpaku, E., & Sosu, R.N. (2025). MTCR-AE: A multiscale temporal convolutional recurrent autoencoder for unsupervised malicious network traffic detection. *Computer Networks*, 261, article number 111147. doi: 10.1016/j.comnet.2025.111147.
[5] Akif, M.A., Butun, I., Williams, A., & Mahgoub, I. (2025). Hybrid machine learning models for intrusion detection in IoT: Leveraging a real-world IoT dataset. *ArXiv*, 2502, article number 12382. doi: 10.48550/arXiv.2502.12382.
[6] Akuthota, U.C., & Bhargava, L. (2025). The role of machine and deep learning in modern intrusion detection systems: A comprehensive review. *Computers and Electrical Engineering*, 124, article number 110318. doi: 10.1016/j.compeleceng.2025.110318.
[7] AL-Husseini, H., Hosseini, M.M., Yousofi, A., & Alazzawi, M.A. (2024). Whale optimization algorithm-enhanced long short-term memory classifier with novel wrapped feature selection for intrusion detection. *Journal of Sensor and Actuator Networks*, 13(6), article number 73. doi: 10.3390/jsan13060073.
[8] Anser, O., François, J., & Chrisment, I. (2024). Automated machine learning configuration to learn intrusion detectors on attack-free datasets. In *Proceedings of the 49th IEEE conference on local computer networks* (pp. 1-7). Normandy: IEEE. doi: 10.1109/LCN60385.2024.10639690.
[9] Canadian Institute for Cybersecurity Intrusion Detection System. (2017). Retrieved from https://www.unb.ca/cic/datasets/ids-2017.html.
[10] Chychkarov, Y., Zinchenko, O., Bondarchuk, A., & Aseeva, L. (2023). Detection of network intrusions using machine learning algorithms and fuzzy logic. *Electronic Professional Scientific Journal "Cybersecurity: Education, Science, Technique"*, 3(19), 209-225. doi: 10.28925/2663-4023.2023.19.209225.
[11] Cohen, S., Goldshlager, N., Shapira, B., & Rokach, L. (2023). TTANAD: Test-time augmentation for network anomaly detection. *Entropy*, 25(5), article number 820. doi: 10.3390/e25050820.
[12] Dairi, A., Harrou, F., Bouyeddou, B., Senouci, S.-M., & Sun, Y. (2023). Semi-supervised deep learning-driven anomaly detection schemes for cyber-attack detection in smart grids. In H.H. Alhelou, N. Hatziargyriou & Z.Y. Dong (Eds.), *Power systems cybersecurity: Methods, concepts, and best practices* (pp. 265-295). Cham: Springer. doi: 10.1007/978-3-031-20360-2_11.
[13] Diana, L., Dini, P., & Paolini, D. (2025). Overview on intrusion detection systems for computers networking security. *Computers*, 14(3), article number 87. doi: 10.3390/computers14030087.
[14] Firdaus, G.M., & Suryani, V. (2024). Enhanced pruning process with DBSCAN for attack detection. In *Proceedings of the 2024 1st international conference on cyber security and computing* (pp. 113-118). Melaka: IEEE. doi: 10.1109/CyberComp60759.2024.10913747.
[15] Gutierrez-Portela, F., Almenares Mendoza, F., & Calderon-Benavides, L. (2024). Evaluation of the performance of unsupervised learning algorithms for intrusion detection in unbalanced data environments. *IEEE Access*, 12, 190134-190157. doi: 10.1109/ACCESS.2024.3516615.
[16] Haidur, H.I., Gakhov, S.O., & Bryhynets, A.A. (2023). Detection of network anomalies with neural networks algorithms. *Telecommunication and Information Technologies*, 78(1), 61-73. doi: 10.31673/2412-4338.2023.010416.
[17] Hassan, T.J., Ramchandra, A.R., & Ranganathan, P. (2024). Modeling and evaluation of false data injection attacks (FDIA) in DER inverters. In *Proceedings of the 56th North American power symposium* (pp. 1-7). El Paso: IEEE. doi: 10.1109/NAPS61145.2024.10741736.
[18] He, H., Liu, K., Zhang, L., Xu, K., Li, J., & Ren, J. (2025). TE-PADN: A poisoning attack defense model based on temporal margin samples. *Big Data Research*, 40, article number 100528. doi: 10.1016/j.bdr.2025.100528.
[19] Ilin, D., & Starinskyi, I. (2023). Mathematical model of an intrusion detection system using a neural network based on auto-encoders. *Modern Information Technologies in the Sphere of Security and Defence*, 47(2), 113-118. doi: 10.33099/2311-7249/2023-47-2-113-118.
[20] Iturbe-Araya, J.I., & Rifà-Pous, H. (2025). Enhancing unsupervised anomaly-based cyberattacks detection in smart homes through hyperparameter optimization. *International Journal of Information Security*, 24, article number 45. doi: 10.1007/s10207-024-00961-6.
[21] Loo, F.-N., Chen, T.-S., Tsai, C.-W., Lin, J., & Yang, C.-W. (2024). Feasibility study of semi-supervised intrusion detection systems for industrial control systems. In *Proceedings of the 10th international conference on applied system innovation* (pp. 166-168). Kyoto: IEEE. doi: 10.1109/ICASI60819.2024.10547975.
[22] Lu, K.-C., Liu, I.-H., Liu, Z.-C., & Li, J.-S. (2024). Common criteria for security evaluation and malicious intrusion detection mechanism of dam supervisory control and data acquisition system. *IET Networks*, 13(5-6), 546-559. doi: 10.1049/ntw2.12127.
[23] Meidan, Y., Avraham, D., Libhaber, H., & Shabtai, A. (2023). CADeSH: Collaborative anomaly detection for smart homes. *IEEE Internet of Things Journal*, 10(10), 8514-8532. doi: 10.1109/JIOT.2022.3194813.
[24] Nikitenko, A., & Bashkov, Y. (2024). Construction of a network intrusion detection system based on a convolutional neural network and a bidirectional gated recurrent unit with attention mechanism. *Eastern-European Journal of Enterprise Technologies*, 3(9), 6-15. doi: 10.15587/1729-4061.2024.305685.

[25] Rajathi, C., & Rukmani, P. (2025). Hybrid learning model for intrusion detection system: A combination of parametric and non-parametric classifiers. *Alexandria Engineering Journal*, 112, 384-396. doi: 10.1016/j.aej.2024.10.101.
[26] Russell-Gilbert, A., Mittal, S., Rahimi, S., Seale, M., Jabour, J., Arnold, T., & Church, J. (2025). RAAD-LLM: Adaptive anomaly detection using LLMs and RAG integration. *ArXiv*, 2503, article number 02800. doi: 10.48550/arXiv.2503.02800.
[27] Sharma, N., Arora, B., Ziyad, S., Singh, P.K., & Singh, Y. (2024). A holistic review and performance evaluation of unsupervised learning methods for network anomaly detection. *International Journal on Smart Sensing and Intelligent Systems*, 17(1), article number 16. doi: 10.2478/ijssis-2024-0016.
[28] Shukla, A.K., Srivastav, S., Kumar, S., & Muhuri, P.K. (2023). UInDeSI4.0: An efficient unsupervised intrusion detection system for network traffic flow in Industry 4.0 ecosystem. *Engineering Applications of Artificial Intelligence*, 120, article number 105848. doi: 10.1016/j.engappai.2023.105848.
[29] Srivastav, S., Shukla, A.K., Kumar, S., & Muhuri, P.K. (2025). HYRIDE: HYbrid and Robust Intrusion DEtection approach for enhancing cybersecurity in Industry 4.0. *Internet of Things*, 30, article number 101492. doi: 10.1016/j.iot.2025.101492.
[30] Tsai, C.-W., Lu, C.-T., Li, C.-H., & Zhang, S.-W. (2024). An effective ensemble electricity theft detection algorithm for smart grid. *IET Networks*, 13(5-6), 471-485 doi: 10.1049/ntw2.12132.
[31] Uddin, M.A., Aryal, S., Bouadjenek, M.R., Al-Hawawreh, M., & Talukder, M.A. (2024). UsfAD based effective unknown attack detection focused IDS framework. *Scientific Reports*, 14, article number 29103. doi: 10.1038/s41598-024-80021-0.
[32] Zoppi, T., Ceccarelli, A., Puccetti, T., & Bondavalli, A. (2023). Which algorithm can detect unknown attacks? Comparison of supervised, unsupervised and meta-learning algorithms for intrusion detection. *Computers and Security*, 127, article number 103107. doi: 10.1016/j.cose.2023.103107.

Виявлення вторгнень у телекомунікаційних мережах за допомогою аналізу мережевого трафіку

Андрій Рій

Аспірант
Національний університет «Львівська політехніка»
79000, вул. Степана Бандери, 12, м. Львів, Україна
https://orcid.org/0009-0005-0252-1533

Мар'ян Кирик

Доктор технічних наук, професор
Національний університет «Львівська політехніка»
79000, вул. Степана Бандери, 12, м. Львів, Україна
https://orcid.org/0000-0001-9156-9347

Анотація. Зростання кіберзагроз і обсягів мережевого трафіку підкреслили потребу в ефективних методах виявлення аномалій. Метою дослідження була розробка гібридної моделі виявлення аномалій у телекомунікаційних мережах із підвищеною точністю на основі порівняльного аналізу безконтрольних алгоритмів. У ході роботи було проведено порівняння чотирьох основних алгоритмів: Isolation Forest, Local Outlier Factor (LOF), One-Class Support Vector Machine (SVM) та Elliptic Envelope, за допомогою симульованих даних мережевого трафіку. Методи аналізу включали нормалізацію даних за допомогою Z-score та Min-Max Scaling для усунення масштабних відмінностей між ознаками. Для підвищення точності виявлення аномалій було застосовано ансамбль із 100 дерев рішень. У результаті було виявлено, що Isolation Forest забезпечував найбільшу точність виявлення аномалій (F1-Score = 85,8 %) і високу швидкість обробки (час обробки на 10000 записів у 2,5 сек), що було важливим для реальних умов телекомунікаційних мереж з великими обсягами даних. LOF показав високу точність при виявленні локальних аномалій, але з більшою обчислювальною складністю. Алгоритм One-Class SVM виявляв глобальні аномалії, але демонстрував нижчу точність для локальних. Elliptic Envelope мав обмежену ефективність через припущення про нормальний розподіл даних. Додатково було розроблено комплексну модель, яка поєднувала переваги Isolation Forest, LOF та Density-Based Spatial Clustering of Applications with Noise, що дозволило підвищити F1-Score до 88 % та перевищити результати окремих моделей. Гібридна модель показала адаптивність до багатомодальних розподілів та ефективність у виявленні локальних аномалій. Практична значимість полягала в підвищенні точності й стійкості систем мережевої безпеки та формуванні основ для адаптивних алгоритмів реального часу, що дозволило фахівцям з кібербезпеки та телекомунікацій ефективніше моніторити та захищати мережі від загроз

Ключові слова: виявлення аномалій; оцінка продуктивності; масштабованість; точність; безконтрольні алгоритми; гібридна модель; кластеризація